

Восстановление гладких монотонных функций

Рассматривается задача восстановления монотонной функции одного переменного по ее (точным или приближенным) значениям в некоторых точках. Предложен вариант четкой постановки функции, описаны методы ее точного и приближенного решения с помощью электронных таблиц. Приведены примеры, когда существующие методы восстановления не обеспечивают монотонности восстановленной функции.

Ключевые слова: функции одного переменного, гладкость, монотонность, восстановление, случайные ошибки.

1. Введение.

Восстановление гладких монотонных функций

На практике нередко возникают задачи восстановления некоторой функции по ее значениям в отдельных точках. В вычислительной математике их называют задачами интерполяции. Типичным экономическим примером является следующий. Известна стоимость прокатных станов годовой производительностью 200, 300 и 1000 тыс. т по состоянию на некоторую дату. Требуется оценить стоимость аналогичного прокатного стана годовой производительностью 700 тыс. т по состоянию на ту же дату.

При проведении массовой оценки подобную интерполяцию приходится производить многократно, в связи с чем необходимы алгоритмы, позволяющие «автоматизировать» этот процесс. К сожалению, известные алгоритмы не гарантируют получение экономически содержательных результатов, например, зависимость стоимости объекта от его мощности оказывается немонотонной. Положение существенно усложняется, если исходная информация о значениях функции является неточной. Предлагаемая статья имеет целью дать вариант четкой постановки задач восстановления монотонной функции одного переменного $y = f(x)$ по ее (точным или приближенным) значениям и описать методы их решения.

2. Базовая ситуация

Начнем с рассмотрения ситуации, когда исходной информацией являются наблюдаемые *точные* значения функции в отдельных точках:

$$y_i = f(x_i); (i = 1, 2, \dots, n; x_1 < \dots < x_n). \quad (1)$$

Если заранее известно, что искомая зависимость — многочлен соответствующей степени, то решение задачи дается известной формулой Ньютона. Если вид зависимости известен с точностью до векторного параметра p , т. е.

$$f(x) = F(x, p),$$

где F — известная функция, то задача сводится к определению этого вектора параметров по исходной информации. Однако в экономических приложениях такое (параметрическое) представление восстанавливаемых функций встречается довольно редко.

В математической литературе описаны решения ряда задач оптимальной интерполяции, т. е. построения алгоритмов, дающих наибольшую точность при восстановлении функций некоторых классов, например, таких, у которых вторая производная по абсолютной величине не превосходит заданной константы (Магарил-Ильяев, Тихомиров, 2003, раздел 6.5).

К сожалению, подобные классы функций практически не встречаются в экономических и некоторых технических исследованиях (о связанных с этим проблемах см. (Клейнер, Смоляк, 2002)). Поэтому на практике восстановление функций обычно осуществляется иными методами.

Чаше всего интерполяция осуществляется с помощью «защитой» в электронные таблицы программы построения графиков, где по желанию пользователя заданные точки соединяются сплайновой линией. Между тем задолго до появления компьютеров инженеры применяли довольно простой способ восстановления, соединяя на графике точки (x_i, y_i) плавной кривой посредством гибкой линейки. Анализ соответствующей процедуры привел в свое время к созданию теории сплайнов. «Инженерный» сплайн при этом являлся решением следующей экстремальной задачи:

$$\int_{-\infty}^{\infty} [f''(x)]^2 dx \Rightarrow \min \quad (2)$$

при ограничениях (1). Действительно, можно показать, что условие (2) примерно характеризует положение деформированной оси гибкой балки, закрепленной в точках (x_i, y_i) .

Аналитические формулы, дающие решение задачи, изложены, например, в (Алберг и др., 1973). В частности, на каждом из отрезков (x_i, x_{i+1}) функция $f(x)$ оказывается кубической параболой (в связи с чем такие функции и именуются кубическими сплайнами), а вне отрезка $[x_1, x_n]$ — линейной.

В книге (Клейнер, Смоляк, 2002) критерию (2) дается «психологическое» объяснение: из нескольких возможных линий, соединяющих данные точки, человек старается выбрать «наиболее гладкую (плавную)», и функционал (2) как раз и характеризует степень «гладкости (плавности)» кривой, не допуская ни изломов, ни резких изменений угла наклона касательной.

Между тем при восстановлении эконометрических зависимостей часто приходится учитывать не только то, что они должны быть достаточно «гладкими», но и некоторые другие их свойства «качественного» характера. Например, из экономических соображений нередко бывает известно, что искомая функция является монотонной, для определенности — монотонно возрастающей. Оказывается, ни электронные таблицы, ни «обычные» кубические сплайны не обеспечивают во всех случаях монотонность восстанавливаемой функции. Желющие могут это проверить на примерах. Условия, при которых монотонность все же обеспечивается, исследовались в ряде работ, например, в (Мирошниченко, 1990; Волков, 2001), но экономического смысла такие условия не имеют.

Казалось бы, решение здесь можно получить, решая задачу (2) при ограничениях (1) и дополнительном ограничении на неотрицательность производной $f'(x)$. Увы, решение такой

задачи в общем случае оказывается следующим: отрезок $[x_1, x_n]$ разбивается на некоторое количество других отрезков, на одних $f(x)$ будет кубической параболой, а на других — константой. При этом точное расположение концов таких отрезков простыми аналитическими формулами не задается. Полученное решение, как правило, не отвечает экономическому содержанию восстанавливаемой функции.

Есть и иной подход к построению монотонных сплайнов, например, см. (Fritsch, Carlson, 1980). Здесь вначале определяются приближенные значения производной восстанавливаемой функции в $f'(x_i)$, а затем на каждом из отрезков (x_i, x_{i+1}) эта функция восстанавливается многочленом третьей степени, имеющим в концах отрезка те же значения $f(x_i), f(x_{i+1}), f'(x_i)$ и $f'(x_{i+1})$. Однако восстановленная таким способом функция также не всегда оказывается монотонной. К тому же какими-то «экстремальными» свойствами она не обладает.

Существуют десятки тысяч работ, посвященных разного вида сплайнам и их применению. Математики преимущественно рассматривают сплайны как метод приближения функций. Поэтому основное внимание они уделяют оценке точности приближения сплайнами функций различных классов, когда количество n наблюдаемых значений функции неограниченно возрастает. Однако там, где возникает необходимость в восстановлении функциональных зависимостей между экономическими показателями, обычно неизвестен ни их вид, ни класс функций, к которому они принадлежат, а объем исходной информации невелик и, как правило, не может быть увеличен. На этом основании вопрос о «точности восстановления» здесь явно отходит на второй план, а «правдоподобность» восстановленной функции в гораздо большей степени определяется ее «гладкостью» и монотонностью.

Принципиально иной подход к восстановлению функций с определенными свойствами предложен в монографии (Клейнер, Смоляк, 2002), где был выдвинут принцип максимальной согласованности. В соответствии с ним критерий «качества восстановления» должен отражать согласованность восстановленной функции со *всей* имеющейся информацией (а не только с данными о ее значениях в заданных точках). В нашем случае такой критерий должен отразить как гладкость функции, так и ее монотонность.

На основании этого более естественным представляется изменить критерий (2) так, чтобы он, обеспечивая достаточную «гладкость» восстановленной функции, одновременно не допускал ни смены знаков ее производной, ни существования «отрезков постоянства». Необходимость использования критериев такого типа была обоснована в (Клейнер, Смоляк, 2002), хотя конкретный предложенный там критерий

$$\int_{-\infty}^{\infty} \left[\frac{f''(x)}{f'(x)} \right]^2 dx \Rightarrow \min$$

оказался неудобным для практического использования. Однако при небольшой модификации указанного критерия можно получить задачу, допускающую аналитическое исследование и решаемую численно на персональных компьютерах¹. Ее постановка выглядит так.

Даны n точек (x_i, y_i) таких, что $x_1 < \dots < x_n$, $y_1 < \dots < y_n$. Найдём функцию $f(x)$, график которой проходит через данные точки, и для которой

¹ Предлагаемая модификация, скорее всего, не является единственно возможной.

$$\int_{-\infty}^{\infty} \frac{[f''(x)]^2}{f'(x)} dx = 4 \int_{-\infty}^{\infty} \left[\left[\sqrt{f'(x)} \right]' \right]^2 dx \Rightarrow \min.$$

Легко видеть, что решением этой задачи будет строго монотонная функция. Ее можно найти следующим путем. Введем в рассмотрение функцию $u(x) = \sqrt{f'(x)}$. Тогда

$$f'(x) = u^2(x), f''(x) = 2u(x)u'(x)$$

и

$$f(x_i) = f(x_{i-1}) + \int_{x_{i-1}}^{x_i} u^2(x) dx, \quad (3)$$

$$\int_{-\infty}^{\infty} \frac{[f''(x)]^2}{f'(x)} dx = 4 \int_{-\infty}^{\infty} u'^2(x) dx = 4H(u). \quad (4)$$

Таким образом, возникает задача найти функцию $u(x)$, дающую минимум функционалу $H(u)$ при ограничениях:

$$\int_{x_i}^{x_{i+1}} u^2(x) dx = y_{i+1} - y_i, \quad (i=1, \dots, n-1). \quad (5)$$

Обозначив множители Лагранжа для уравнений (5) через r_i , сведем эту задачу к задаче безусловной минимизации:

$$\int_{-\infty}^{\infty} u'^2(x) dx + \sum_{i=1}^{n-1} r_i \left[\int_{x_i}^{x_{i+1}} u^2(x) dx - (y_{i+1} - y_i) \right] \Rightarrow \min.$$

Решение последней задачи получим известными методами вариационного исчисления, оно оказывается достаточно простым: на отрезке (x_i, x_{i+1}) функция $u(x)$ будет удовлетворять уравнению:

$$u''(x) - r_i u(x) = 0,$$

вне отрезка $[x_1, x_n]$ —

$$u'(x) = 0,$$

а в точках x_1, \dots, x_n правые пределы самой функции $u(x)$ и ее производной будут совпадать с левыми (это так называемые условия «гладкой склейки»):

$$u'(x_1 + 0) = 0; \quad u'(x_i - 0) = u'(x_i + 0); \quad (i=2, \dots, n-1); \quad u'(x_n - 0) = 0.$$

Для получения расчетных формул заметим следующее. Как известно, общее решение дифференциального уравнения $u''(x) - r_i u(x) = 0$ является либо линейной функцией (при $r_i = 0$), либо линейной комбинацией тригонометрических (при $r_i < 0$) или гиперболических

(при $r_i > 0$) функций². Через эти же функции выражаются и интегралы от $u^2(x)$. Чтобы получить обозримые расчетные формулы, полезно ввести следующие обозначения:

$$s_i = \sqrt{|r_i|}, \quad D_i = x_{i+1} - x_i, \quad u_i = u(x_i);$$

$$w_i(x) = \begin{cases} \frac{\operatorname{sh}(s_i x)}{s_i} & \text{при } r_i > 0; \\ x & \text{при } r_i = 0; \\ \frac{\sin(s_i x)}{s_i} & \text{при } r_i < 0. \end{cases} \quad v_i(x) = w'_i(x) = \begin{cases} \operatorname{ch}(s_i x) & \text{при } r_i \geq 0; \\ \cos(s_i x) & \text{при } r_i \leq 0; \end{cases} \quad (6)$$

$$A_i(x) = \int_0^x w_i^2(t) dt = \frac{w_i(2x) - 2x}{4r_i} = \frac{w_i(x)v_i(x) - x}{2r_i};$$

$$B_i(x) = \int_0^x w_i(t)w_i(D_i - t) dt = \frac{w_i(D_i - 2x) - w_i(D_i) + 2xv_i(D_i)}{4r_i}.$$

Тогда на отрезке (x_i, x_{i+1}) будет:

$$u(x) = \frac{u_i w_i(x_{i+1} - x) + u_{i+1} w_i(x - x_i)}{w_i(D_i)}. \quad (7)$$

Поэтому на данном отрезке искомая функция определяется формулой:

$$f(x) = y_i + \int_{x_i}^x u^2(t) dt =$$

$$= y_i + \frac{u_{i+1}^2 A_i(x - x_i) + 2u_i u_{i+1} B_i(x - x_i) + u_i^2 [A_i(D_i) - A_i(x_{i+1} - x)]}{w_i^2(D_i)}. \quad (8)$$

Дифференцируя формулу (7), получаем:

$$u'(x) = \frac{u_{i+1} v_i(x - x_i) - u_i v_i(x_{i+1} - x)}{w_i(D_i)}. \quad (9)$$

Поскольку $v_i(0) = 1$, условия «гладкой склейки» принимают вид:

$$\frac{u_2 - u_1 v_1(D_1)}{w_1(D_1)} = 0; \quad (10)$$

² Для читателей, не знакомых с высшей математикой, напомним, что гиперболические синус и косинус определяются формулами: $\operatorname{sh} z = (e^z - e^{-z})/2$, $\operatorname{ch} z = (e^z + e^{-z})/2$. В электронных таблицах они вычисляются с помощью встроенных функций \sinh и \cosh .

$$\frac{u_i v_{i-1}(D_{i-1}) - u_{i-1}}{w_{i-1}(D_{i-1})} = \frac{u_{i+1} - u_i v_i(D_i)}{w_i(D_i)}, \quad (i = 2, \dots, n-1); \quad (11)$$

$$\frac{u_n v_{n-1}(D_{n-1}) - u_{n-1}}{w_{n-1}(D_{n-1})} = 0. \quad (12)$$

Кроме того, величины u_i и r_i должны удовлетворять условиям (5). Подставляя в (8) $x = x_{i+1}$, эти условия можно записать следующим образом:

$$\frac{(u_{i+1}^2 + u_i^2)A_i(D_i) + 2u_i u_{i+1}B_i(D_i)}{w_i^2(D_i)} = y_{i+1} - y_i, \quad (i = 1, \dots, n-1). \quad (13)$$

Мы получили систему $2n - 1$ уравнений (10)–(13) с $2n - 1$ неизвестными u_i и r_i . Решив данную систему³, значения искомой функции на отрезке $[x_1, x_n]$ можно найти по формуле (8).

Проведя касательные к графику этой функции в крайних точках, получим выражения для $f(x)$ вне отрезка $[x_1, x_n]$:

$$f(x) = y_1 + u_1^2(x - x_1), \quad (x \leq x_1); \quad f(x) = y_n + u_n^2(x - x_n), \quad (x \geq x_n).$$

Пример 1. На рис. 1 сплошной линией показан результат восстановления монотонной функции по следующим исходным данным (соответствующие точки на графике выделены ромбиками).

x_i	1	1.8	2.7	3.2	4.2	5.2	6
y_i	1	1.4	2.7	8	9	26	30

Значения u_i и r_i были найдены с помощью опции «Поиск решения» в Excel и оказались следующими.

i	1	2	3	4	5	6	7
u_i	0.79	0.53	2.71	2.19	1.91	3.59	1.51
r_i	–1.07	9.44	–11.09	20.54	–3.61	3.58	0.00

Пунктиром на рисунке показана та же функция, восстановленная с помощью Excel. Следует обратить внимание на то, что она оказалась немонотонной.

³ Здесь и далее мы не будем уделять особого внимания численным методам решения систем нелинейных уравнений, поскольку разработка таких методов является предметом исследований вычислительной математики, а не эконометрики. Приводимые ниже примеры решались с использованием опции «Поиск решения» в Excel. Ряд трудностей, возникающих при применении такого метода решения, обсудим далее.

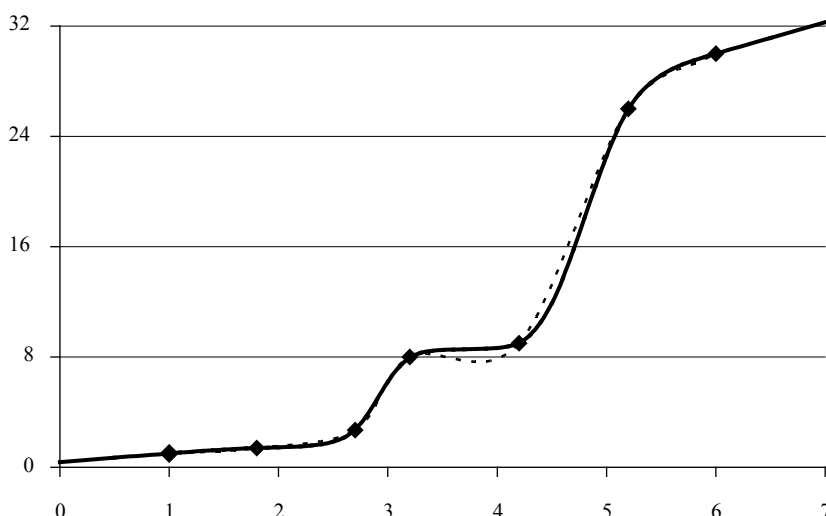


Рис. 1. Результаты восстановления функции в примере 1

3. Особые краевые условия

В рассмотренной выше базовой ситуации восстановленная функция за пределами отрезка $[x_1, x_n]$ — линейная и неограниченно возрастает по абсолютной величине при $|x| \rightarrow +\infty$. Однако часто бывает известно, что восстанавливаемая функция при всех значениях аргумента, например, положительна и/или не превосходит 1. Ни кубические сплайны, ни ранее описанный метод не позволяют получить такого результата. В связи с этим естественно рассмотреть задачу восстановления монотонной функции при дополнительном ограничении $f(-\infty) = 0$, подразумевая, что $y_1 > 0$.

Для ее решения, как и раньше, будем минимизировать функционал (4) при ограничениях $f(x_i) = y_i$. Но здесь

$$f(x_i) = \int_{-\infty}^{x_i} f'(x) dx = \int_{-\infty}^{x_i} u^2(x) dx,$$

поэтому указанные ограничения принимают вид:

$$\int_{-\infty}^{x_1} u^2(x) dx = y_1, \int_{x_i}^{x_{i+1}} u^2(x) dx = y_{i+1} - y_i, \quad (i = 1, \dots, n-1).$$

Полученная задача эквивалентна задаче безусловной минимизации:

$$\int_{-\infty}^{\infty} u^2(x) dx + r_0 \int_{-\infty}^{x_1} u^2(x) dx + \sum_{i=1}^{n-1} r_i \left[\int_{x_i}^{x_{i+1}} u^2(x) dx - (y_{i+1} - y_i) \right] \Rightarrow \min.$$

Решение этой задачи получаем известными методами вариационного исчисления. Оказывается, что функция $u(x)$ удовлетворяет следующим условиям: на луче $x < x_1$ должно быть

$$u''(x) - r_0 u(x) = 0,$$

на каждом из отрезков (x_i, x_{i+1})

$$u''(x) - r_i u(x) = 0,$$

а на луче $x > x_n$

$$u'(x) = 0,$$

причем в точках x_1, \dots, x_n должны выполняться условия «гладкой склейки».

Вначале рассмотрим луч $x < x_1$. Здесь $u''(x) - r_0 u(x) = 0$. Пусть $s_0 = \sqrt{|r_0|}$. Тогда общее решение этого уравнения можно записать в следующем виде:

$$u(x) = \begin{cases} C_1 e^{s_0 x} + C_2 e^{-s_0 x} & \text{при } r_0 > 0; \\ C_1 x + C_2 & \text{при } r_0 = 0; \\ C_1 \sin(s_0 x) + C_2 \cos(s_0 x) & \text{при } r_0 < 0. \end{cases}$$

Заметим теперь, что $\int_{-\infty}^{x_1} u^2(x) dx = y_1$, поэтому стоящий слева интеграл — конечный и не равен нулю. Легко видеть, что это условие не выполняется ни при каких значениях C_1 и C_2 , если только $r_0 \leq 0$.

Если же $r_0 > 0$, то это условие будет выполняться только при $C_1 = 0$. Но тогда $u(x) = C_2 e^{s_0 x}$ и, в частности, $u_1 = u(x_1) = C_2 e^{s_0 x_1}$.

В таком случае $C_2 = u_1 e^{-s_0 x_1}$, и значит $u(x) = u_1 e^{s_0(x-x_1)}$.

Поэтому $y_1 = \int_{-\infty}^{x_1} u^2(x) dx = \frac{u_1^2}{2s_0}$, так что $s_0 = \frac{u_1^2}{2y_1}$. Отсюда при $x < x_1$ имеем:

$$\begin{aligned} f(x) &= \int_{-\infty}^x u^2(t) dt = \frac{u_1^2}{2s_0} e^{2s_0(x-x_1)} = y_1 e^{2s_0(x-x_1)} = y_1 \exp\left[\frac{u_1^2}{y_1}(x-x_1)\right]; \\ \int_{-\infty}^{x_1} u'^2(t) dt &= s_0^2 \int_{-\infty}^{x_1} u^2(t) dt = s_0^2 y_1 = \frac{u_1^4}{4y_1}. \end{aligned} \quad (14)$$

Отметим, что $u'(x) = s_0 u(x)$, вследствие этого

$$u'_1 = s_0 u_1 = \frac{u_1^3}{2y_1}.$$

Отсюда и из условия «гладкой склейки» с учетом (9) получаем:

$$\frac{u_2 - u_1 v_1(D_1)}{w_1(D_1)} = \frac{u_1^3}{2y_1}. \quad (15)$$

Таким образом, в данной ситуации требуется решить ту же систему уравнений, что и в базовой, с той лишь разницей, что уравнение (9) должно быть заменено на (15).

Аналогично рассматривается ситуация с дополнительным ограничением $f(\infty) = c$ или с двумя ограничениями на $f(-\infty)$ и $f(\infty)$. Такие ситуации возникают, например, при восстановлении функции распределения случайной величины по сгруппированным ею наблюдаемым значениям, если, например, первая из групп включает все значения случайной величины, меньшие некоторой границы, а последняя группа — все значения случайной величины, большие некоторой другой границы (см. ниже).

4. Восстановление монотонной функции при наличии ошибок наблюдений

Ранее, при рассмотрении задачи восстановления функции по ее значениям, мы предполагали, что эти значения известны точно. Между тем на практике часто встречается ситуация, когда наблюдаемые значения функции по тем или иным причинам отклоняются от ее истинных значений. Такие отклонения обычно объясняются влиянием случайных факторов. Приведем одну из возможных моделей восстановления монотонной функции в подобной ситуации.

Исходными данными в этой модели являются наблюдаемые значения (y_i) некоторой монотонной функции в точках x_i , которые могут отличаться от истинных значений этой функции (f_i). Поэтому, в частности, наблюдаемые значения функции могут меняться уже не монотонно (т. е. неравенства $y_i > y_{i-1}$ могут не выполняться). Принимается, что отклонения $y_i - f_i$ — случайные величины с нулевым средним значением и известной дисперсией S .

Последнее предположение требует определенных оправданий.

С одной стороны, в условиях, когда дисперсия отклонений неизвестна, решение задачи оказывается тривиальным: восстанавливаемую функцию можно считать тождественно равной нулю (наиболее «гладкая» из всех известных функций), а все фактически наблюдаемые значения функции рассматривать как случайные ошибки. И доказать, что это не так, невозможно.

С другой стороны, когда экономист говорит, что он *не знает* дисперсию отклонений, это не означает, что он *ничего не знает* об этой дисперсии. На самом деле *общее представление о величине этой дисперсии имеется всегда*. Типичным примером являются задачи оценки стоимости имущества по данным о ценах рыночных сделок.

Оценщики знают, что обычно аналогичные (идентичные по основным характеристикам) объекты продаются на рынке в одно и то же время по разным ценам. Расхождение в ценах несет информацию о величине случайных отклонений цен сделок от рыночной стоимости (на разных рынках дисперсия таких отклонений может быть различной). Иными словами, дисперсия цен сделок с аналогичными объектами — объективная характеристика рынка, она поддается измерению, и оценщики при анализе рынков всегда могут сказать, что на одном рынке колебания цен сделок составляют 5–10%, а на другом — 30–50%.

Но имеющаяся информация не позволяет указать величину дисперсии *точно*. Однако представленный ниже метод, как будет показано далее, оказывается применимым и тогда, когда она известна приближенно.

Основываясь на изложенных соображениях, будем искать наиболее «гладкую» в указанном выше смысле монотонную функцию, отклонения истинных значений которой от наблюдаемых имеют заданную дисперсию. Для этого нужно минимизировать функционал (4) при ограничении

$$\frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2 = S. \quad (16)$$

Заметим, что при нормально распределенных случайных отклонениях это условие означает, что истинная дисперсия отклонений совпадает с ее наиболее правдоподобным значением — фактической (выборочной) дисперсией отклонений.

В силу (3) ограничение (16) можно записать иначе:

$$\sum_{i=1}^n \left(f_i + \int_{x_1}^{x_i} u^2(x) dx - y_i \right)^2 = nS.$$

Обозначив через $r/2$ множитель Лагранжа, отвечающий этому ограничению, решение данной задачи можно свести к задаче безусловной оптимизации:

$$\int_{-\infty}^{\infty} u'^2(x) dx + \frac{r}{2} \left[\sum_{i=1}^n \left(f_i + \int_{x_1}^{x_i} u^2(x) dx - y_i \right)^2 - nS \right] \Rightarrow \min. \quad (17)$$

Для решения задачи обозначим:

$$f_i = f(x_i), \quad r_i = r[(f_i - y_i) + (f_{i+1} - y_{i+1}) + \dots + (f_n - y_n)]. \quad (18)$$

Тогда условиями оптимальности f_i , r_i и $u(x)$ будут:

- 1) $r_1 = 0$;
- 2) $u(x)$ имеет непрерывную производную и на каждом отрезке (x_i, x_{i+1}) удовлетворяет уравнению $u''(x) - r_i u(x) = 0$;
- 3) $u'(x) = 0$ вне отрезка $[x_1, x_n]$;
- 4) в точках x_1, \dots, x_n выполняются условия «гладкой склейки»:

$$u'(x_1 + 0) = 0; \quad u'(x_i - 0) = u'(x_i + 0); \quad (i = 2, \dots, n-1); \quad u'(x_n - 0) = 0.$$

Соответствующая система уравнений будет включать уравнения (10)–(12) без всяких изменений, в уравнениях (13) наблюдаемые значения функции заменятся на истинные:

$$\frac{(u_{i+1}^2 + u_i^2)A_i(D_i) + 2u_i u_{i+1} B_i(D_i)}{w_i^2(D_i)} = f_{i+1} - f_i, \quad (i = 1, \dots, n-1),$$

и, наконец, добавятся дополнительные уравнения, вытекающие из (18):

$$r_i = r[(f_i - y_i) + (f_{i+1} - y_{i+1}) + \dots + (f_n - y_n)], \quad (i = 1, \dots, n); \quad r_1 = 0.$$

Множитель r при этом должен быть таким, чтобы выполнялось условие (16). Для этих целей удобно решать данную задачу при разных значениях r , затем находить соответствующую

щее S по формуле (16) и интерполировать получающуюся зависимость $S(r)$, которая будет иметь следующий вид. Если r очень велико, значения f_i будут получаться близкими к y_i , так что S окажется небольшим. При уменьшении r до нуля функция $f(x)$ постепенно сглаживается, пока не превращается в линейную⁴. Средний квадрат отклонений f_i от y_i будет при этом расти. «Предельной» линейной функции будет отвечать какое-то значение $S = S^*$ и, следовательно, решение рассматриваемой задачи имеет практический смысл, только если $S < S^*$.

Пример 2. Приведем два варианта восстановления монотонной функции по следующим данным.

x_i	0	1	2	3	5
y_i	3	5	14	24	29

В варианте 1 дисперсия S отклонений $y_i - f_i$ равна 0.4, в варианте 2 — 1.0. Решение представлено на рис. 2 (первому варианту соответствует жирная линия). Исходные данные выделены на рисунке ромбиками. Как видим, результаты восстановления при разных значениях S оказываются достаточно близкими.

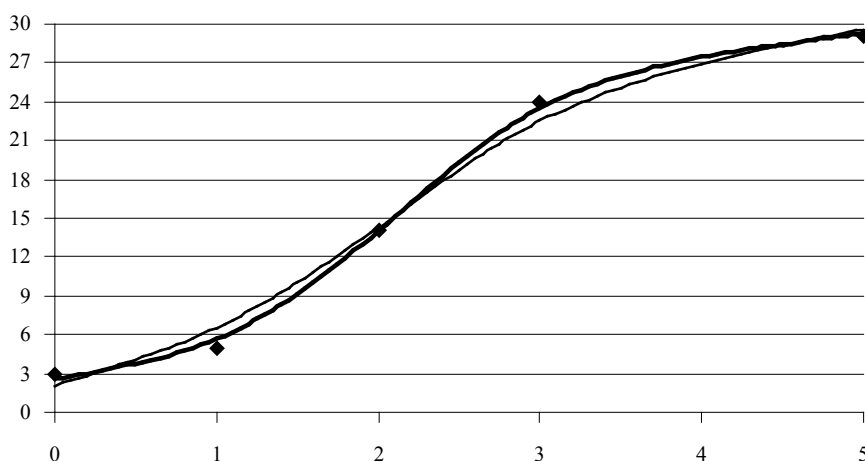


Рис. 2. Восстановленная функция в примере 2

Если точное значение дисперсии отклонений S неизвестно, но известен интервал, в котором она находится, изложенный метод можно применять, приняв в качестве S ее наибольшее возможное значение, если только оно меньше S^* или S^* — в противном случае. Говорить о задаче выбора какого-то «наиболее оптимального» значения S в соответствующем интервале не имеет смысла.

Действительно, допустим, что S лежит в пределах от 2 до 6, причем $S^* > 6$. Тогда любая функция, для которой

⁴ «Предельная» линейная функция устроена просто. Построим уравнение линейной регрессии по точкам (x_i, y_i) : $y = a + bx$. Тогда при положительном b «предельная» линейная функция будет задаваться тем же уравнением, в противном случае — тождественно равна среднему арифметическому из величин y_i .

$$2 \leq \frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2 \leq 6,$$

должна рассматриваться как «приемлемая», согласованная с исходной информацией. Однако из таких функций наиболее «гладкой» будет та, которая построена вышеизложенным методом при $S = 6$. Если же $S^* \leq 6$, то наиболее «гладкой» из «приемлемых» функций окажется указанная выше линейная⁵.

5. Вычислительные проблемы

Численное решение приведенных выше систем уравнений может представлять определенные трудности. Одна из них связана с тем, что при неточном выборе значений r_i уравнения (10)–(12) будут выполняться, только если все $u_i = 0$, тогда как остальные уравнения системы при этом выполняться не будут.

Другая трудность возникает в связи с тем, что при заданном u_i уравнение (13) может иметь два корня u_{i+1} , и априори неясно, какой из них нужно выбрать. Если численное решение осуществляется в Excel с помощью опции «Поиск решения», то корень выбирается автоматически, но этот выбор зависит от выбора начальных значений отыскиваемых переменных, причем при неправильном выборе найти решение системы уравнений не удастся. Поэтому рассмотренные ранее примеры решались при разных вариантах начальных значений переменных, пока не удавалось удовлетворить всем уравнениям системы. Построение более эффективного и «автоматизированного» алгоритма решения системы (10)–(13) требует специального исследования.

В связи с этим представляет интерес другой подход к численному решению рассмотренных задач. Разберем его вначале на примере базовой ситуации.

Разобьем каждый из отрезков (x_i, x_{i+1}) на четное число $2m_i$ маленьких отрезков равной длины, и все полученные точки разбиения, включая все точки x_i , перенумеруем по порядку, начиная с 0, а точку с номером k обозначим z_k .

Пусть

$$u_k = u(z_k), u'_k = u'(z_k), f_k = f(z_k).$$

Предположим, что все величины u_k известны. Тогда значения f_k и критериального функционала $H(u)$ можно приближенно рассчитать, используя известные формулы приближенного интегрирования и дифференцирования.

Действительно, величина f_0 известна точно: $f_0 = y_0$. Другие значения функции f на отрезке $[x_1, x_2]$ находят по следующим рекуррентным формулам приближенного интегрирования:

$$\begin{aligned} f_{2k+1} &= f_{2k} + \int_{z_{2k}}^{z_{2k+1}} u^2(x) dx \approx f_{2k} + (z_{2k+1} - z_{2k}) \frac{5u_{2k}^2 + 8u_{2k+1}^2 - u_{2k+2}^2}{12}; \\ f_{2k+2} &= f_{2k} + \int_{z_{2k}}^{z_{2k+2}} u^2(x) dx \approx f_{2k} + (z_{2k+1} - z_{2k}) \frac{u_{2k}^2 + 4u_{2k+1}^2 + u_{2k+2}^2}{3}. \end{aligned} \quad (19)$$

⁵ Мы не рассматриваем аномальной ситуации, когда $S^* < 2$. Ее возникновение, по нашему мнению, говорит об ошибке в установлении граничных значений S .

Эти формулы позволяют сначала найти $f(x_2) = f(z_{2m_1}) = f_{2m_1}$, затем — значения функции f на отрезке $[x_2, x_3]$ и далее до последней точки x_n .

Значения критериального функционала рассчитаем по следующей приближенной квадратурной формуле:

$$H(u) = \int_{-\infty}^{\infty} u'^2(x) dx = \int_{x_1}^{x_n} u'^2(x) dx = \sum_k \int_{z_{2k}}^{z_{2k+2}} u'^2(x) dx \approx \sum_k \frac{8(u_{2k+1} - u_{2k})^2 + 8(u_{2k+2} - u_{2k+1})^2 - (u_{2k+2} - u_{2k})^2}{6(z_{2k+1} - z_{2k})}. \quad (20)$$

Таким образом, получаем задачу: найти такие u_k , чтобы сумма (20) приняла наименьшее значение при условии, что значения функции f в точках x_2, \dots, x_n совпадают с заданными:

$$f_{2m_1} = y_2, f_{2m_1+2m_2} = y_3, \dots$$

Эту задачу несложно решить в электронных таблицах с помощью опции «Поиск решения». Отметим два важных преимущества такого метода.

Во-первых, он работает практически при любом выборе начальных значений u_k (даже если положить все $u_k = 0$).

Во-вторых, допускает двойной контроль правильности решения. Действительно, по ходу расчетов (ручных или автоматизированных) вычисляется значение целевой функции $H(u)$ так, что всегда можно проверить, что в ходе итерационного подбора величин u_k оно уменьшается.

Кроме того, после завершения итерационного процесса возможно оценить точность полученного решения, а именно — его считают достаточно точным, если правые и левые производные функции $u(x)$ в точках разбиения z_k с достаточной точностью совпадают, а правая производная в точке x_1 и левая производная в точке x_n равны нулю. Чтобы проверить это, необходимо рассчитать значения правой и левой производных в точках разбиения. Для этого можно использовать следующие приближенные формулы:

$$f'(z_k - 0) = \frac{3u_k - 4u_{k-1} + u_{k-2}}{2}; \quad f'(z_k + 0) = \frac{-3u_k + 4u_{k+1} - u_{k+2}}{2}. \quad (21)$$

Оказывается, приемлемой точности добиваются, разбивая отрезки (x_i, x_{i+1}) не более чем на 6–10 частей.

Изложенный выше метод можно применить и при наличии особых краевых условий. Так, в разделе 3 рассмотрена ситуация с дополнительным ограничением $f(-\infty) = 0$.

Здесь в соответствии с (14) выражение для целевой функции изменяется, и для получения решения задачи необходимо только скорректировать квадратурную формулу (20). Поскольку в принятых обозначениях $x_1 = z_0$, $u(x_1) = u_0$, эта формула принимает вид:

$$H(u) = \int_{-\infty}^{\infty} u'^2(x) dx = \int_{-\infty}^{x_1} u'^2(t) dt + \sum_k \int_{z_{2k}}^{z_{2k+2}} u'^2(x) dx \approx \frac{u_0^4}{4y_1} + \sum_k \frac{8(u_{2k+1} - u_{2k})^2 + 8(u_{2k+2} - u_{2k+1})^2 - (u_{2k+2} - u_{2k})^2}{6(z_{2k+1} - z_{2k})}. \quad (22)$$

Значения $f(x)$ при $x < x_1$ при этом находим по формуле (14).

Аналогично учитывается дополнительное ограничение $f(\infty) = c$.

Пример 3. Рассмотрим следующие данные о распределении промышленного оборудования в Российской Федерации по возрасту за 1998 г.

Возраст оборудования	Доля в парке
До 5 лет	0.041
6–10 лет	0.201
11–15 лет	0.256
15–20 лет	0.186
Свыше 20 лет	0.316

В данном случае естественно отыскивать монотонную (кумулятивную) функцию распределения $f(x)$, отражающую долю промышленного оборудования с возрастом, не превышающим x . Очевидно, что она должна удовлетворять условиям: $f(0) = 0$, $f(\infty) = 1$.

На рис. 3 приведен результат восстановления этой функции. Ромбиками на графике выделены точки, отвечающие исходной информации.

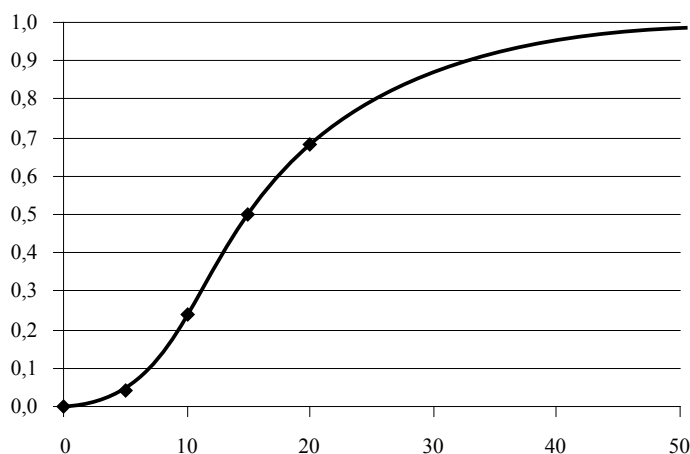


Рис. 3. Восстановленная функция распределения оборудования по возрасту

В ситуации, когда значения функции заданы неточно, к числу неизвестных добавляется еще и f_1 , а требование точного прохождения графика восстанавливаемой функции через заданные точки заменяется ограничением по дисперсии отклонений:

$$(f_0 - y_1)^2 + (f_{2m_1} - y_2)^2 + (f_{2m_1+2m_2} - y_3)^2 + \dots = nS.$$

Пример 4. Восстановим монотонную функцию по следующим исходным данным.

x_i	0	0.7	1	1.6	2	3	3.9	4.3	5
y_i	3.5	5	4.5	6.5	17	23	21	27	25

Среднеквадратичное отклонение наблюдаемых значений функции примем на уровне 2.0. Результат восстановления показан на рис. 4. Там же для сравнения приведен график линейной регрессионной зависимости (пунктирная линия), дающей среднеквадратичное отклонение 3.19 (почти в 1.7 раза больше).

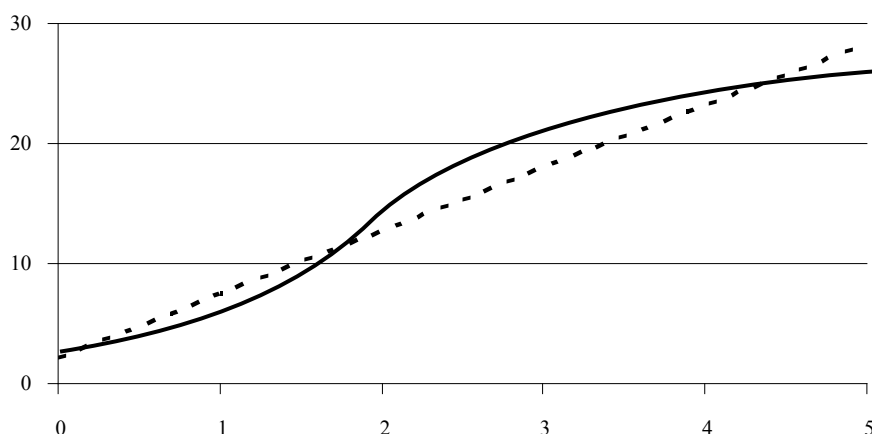


Рис. 4. Результат восстановления функции из примера 4

У подобных задач, как уже отмечалось, могут быть «локальные оптимумы». Избежать их можно следующим способом.

На начальном шаге процедуры рассчитаем уравнение линейной регрессии по точкам (x_i, y_i) : $y = a + bx$. Будем считать, что $b > 0$, так что соответствующая регрессионная зависимость — возрастающая. Средний квадрат отклонений $y_i - a - bx_i$ обозначим через S^* .

Таким образом, для $S = S^*$ решение задачи:

$$f_1 = a + bx_1, \quad u_k = \sqrt{b} \quad \forall k.$$

На следующем шаге немного уменьшим величину S . Соответствующее решение задачи найдем, используя опцию «Поиск решения» в электронных таблицах, взяв полученные на предыдущем шаге значения f_1 и u_k в качестве первого приближения. Далее будем повторять ту же процедуру, постепенно уменьшая значение S , пока не дойдем до требуемого значения. Этот процесс можно довести и до $S = 0$. В этом случае мы получим решение задачи для базовой ситуации.

На рис. 5 представлена зависимость значений критериального функционала (22) от дисперсии отклонений S для исходных данных из примера 2.

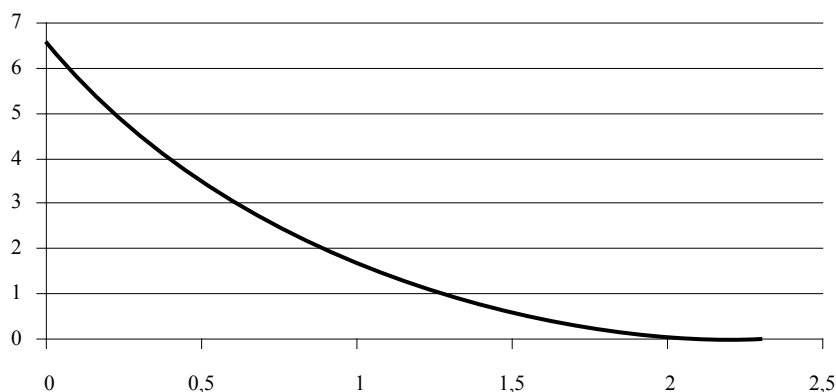


Рис. 5. Зависимость критерия «гладкости» от дисперсии отклонений S

Возможен и иной подход к численному решению рассмотренных задач. Заметим вначале, что:

$$\int_0^{D_i} v_i^2(t) dt = \frac{w_i(D_i)v_i(D_i) + D_i}{2}; \quad \int_0^{D_i} v_i(t)v_i(D_i - t) dt = \frac{w_i(D_i) + D_i v_i(D_i)}{2}.$$

Тогда в силу (9) получим:

$$\int_{x_i}^{x_{i+1}} u'^2(x) dx = \frac{(u_i^2 + u_{i+1}^2)[w_i(D_i)v_i(D_i) + D_i] - 2u_i u_{i+1}[w_i(D_i) + D_i v_i(D_i)]}{2w_i^2(D_i)}.$$

Поэтому в базовой ситуации имеем:

$$\begin{aligned} H(u) &= \int_{-\infty}^{\infty} u'^2(x) dx = \\ &= \sum_i \frac{(u_i^2 + u_{i+1}^2)[w_i(D_i)v_i(D_i) + D_i] - 2u_i u_{i+1}[w_i(D_i) + D_i v_i(D_i)]}{2w_i^2(D_i)}. \end{aligned} \quad (23)$$

В таком случае решение задачи находят, минимизируя функционал (23) по переменным u_i и r_i при ограничениях (13). В правильности найденного решения можно убедиться, проверяя условия «гладкой склейки».

При экспериментальных расчетах в Excel иногда опция «Поиск решения» не могла найти решения задачи. Исправить положение оказалось возможным в несколько шагов.

На *первом шаге* построим какую-то «не слишком изогнутую» кривую, проходящую вблизи заданных точек. Абсциссам x_i на этой кривой будут отвечать какие-то ординаты g_i . Задав точки (x_i, g_i) в качестве исходных, указанным выше методом можно найти отвечающие им значения u_i и r_i . Теперь на *каждом следующем шаге* будем немного изменять значения g_i , приближая их к заданным значениям y_i , и решать задачу для нового набора (x_i, g_i) . При этом значения u_i и r_i , найденные на предыдущем шаге, принимаются в качестве первого приближения при поиске их новых оптимальных значений.

Тот же способ можно использовать и тогда, когда для восстановления функции применяется система уравнений (10) – (13).

6. Заключение

Приведенные примеры показывают, что задача восстановления монотонных функций является актуальной. Помимо рассмотренных в статье ситуаций, предложенные методы могут быть использованы при восстановлении монотонных сепарабельных функций нескольких переменных. Подобные задачи встречаются, например, при восстановлении зависимости стоимости оборудования от его конструктивных параметров.

Однако задача восстановления функции многих переменных, монотонной по каждому из них (например, функции двумерного вероятностного распределения), представляется значительно более трудной. Следует отметить также, что предложенные в статье алгоритмы восстановления монотонных функций, скорее всего, не являются оптимальными в вычислительном отношении и потому нуждаются в совершенствовании.

Список литературы

- Алберг Дж., Нильсен Э., Уолш Дж. (1973). *Теория сплайнов и ее приложения*. М.: Мир.
- Волков Ю. С. (2001). О монотонной интерполяции кубическими сплайнами. *Вычислительные технологии*, 6 (6), 14–24.
- Клейнер Г. Б., Смоляк С. А. (2002). *Эконометрические зависимости: принципы и методы построения*. М.: Наука.
- Магарил-Ильяев Г. Г., Тихомиров В. М. (2003). *Выпуклый анализ и его приложения*. М.: УРСС.
- Мирошниченко В. Л. (1990). Достаточные условия монотонности и выпуклости для интерполяционных кубических сплайнов класса C^2 . *Вычислительные системы: Сб. научных трудов*. Вып. 137 «Приближение сплайнами». АН СССР, Сибирское отделение, Институт математики. 31–57.
- Fritsch F. N., Carlson R. E. (1980). Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17 (2), 238–246.